ForCES Working Group
Internet Draft
Document: draft-ietf-forces-evaluation-00.txt
Expires: June 2004

D. Putzolu (editor)
Intel

December 2003

ForCES Protocol Evaluation Draft

Status of this Memo

Abstract

This document provides an evaluation of the applicability of three proposed approaches for a ForCES protocol: FACT[2], GRMP[3], and Netlink2[4]. A summary of each of the proposed protocols against the ForCES requirements[5] and the ForCES framework[6] is provided. Compliancy of each of the protocols against each requirement is detailed. A conclusion summarizes how each of the protocols fares in the evaluation.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [7].

Table of Contents

1. Introduction....2
2. Protocol Proposals....3
2.1 FACT....4
2.2 GRMP....5
2.3 Netlink2....6
3. Architectural Requirements Compliance Evaluation....8
3.1 FACT....8
3.2 GRMP....8
3.3 Netlink2....10
4. Model Requirements Compliance Evaluation....12
4.1 FACT....12
4.2 GRMP....12
4.3 Netlink2....13
5. Protocol Requirements Compliance Evaluation....13
5.1 Protocol Requirement: Configuration of Modeled Elements...14
5.2 Protocol Requirement: Support for Secure Communication....15
5.3 Protocol Requirement: Scalability....16
5.4 Protocol Requirement: Multihop....17
5.5 Protocol Requirement: Message Priority....18
5.6 Protocol Requirement: Reliability....19
5.7 Protocol Requirement: Interconnect Independence....20
5.8 Protocol Requirement: CE Redundancy or CE Failover....21
5.9 Protocol Requirement: Packet Redirection/Mirroring....22
5.10 Protocol Requirement: Topology Exchange....23
5.11 Protocol Requirement: Dynamic Association....24
5.12 Protocol Requirement: Command Bundling....25
5.13 Protocol Requirement: Asynchronous Event Notification....26
5.14 Protocol Requirement: Query Statistics....26
5.15 Protocol Requirement: Protection Against Denial of Service Attacks....27
5.16 Protocol Requirement Summary Table....28
Security Considerations....29
References....29
Author's Addresses....30

1. Introduction

This document provides an evaluation of the applicability of FACT, GRMP, and Netlink2 as the ForCES protocol. This evaluation provides overviews of the protocols and general statements of applicability based upon the ForCES framework and requirements documents. The format and structure as well as some of the introductory content of this document is based on and taken from a similar document being produced in the MIDCOM working group[8].

The process for protocol evaluation found in this document consists of individuals providing sections evaluating a specific protocol. These sections are incorporated by the editor of the document, and are subject to feedback and changes based on the consensus of the ForCES working group. Some protocols that might be considered as potentially applicable as the ForCES protocol are not evaluated in this document since there where no champions to submit evaluations for them.

Section 2 of this document is a list of the proposed protocols along with background information about each of the protocols.

Section 3 of this document is an evaluation of the proposed protocols against the architectural requirements found in section 5 of the ForCES requirements. The purpose of this section is to determine how well each of the proposed protocols maps to the ForCES architecture.

Section 4 of this document is an evaluation of the proposed protocols against the model requirements found in ForCES requirements. The purpose of this section is to determine how well each of the proposed protocols can be used with FEs that meet the ForCES model requirements.

Section 5 of this document is an item level evaluation of the proposed protocols against the protocol requirements found in the ForCES requirements. The purpose of this section is to determine how well each of the proposed protocols satisfies each of the protocol requirements.

Section 6 summarizes the evaluation, and includes a table with a breakdown for each of the protocols versus the requirements. The following categories of compliance are used: Fully met, partially met through the use of extensions, partially met through other changes to the protocol, or not met. This summary is not a conclusive statement of the suitability of the protocols, but rather to provide information to be considered as input into the overall protocol decision process.

2. Protocol Proposals

The following protocols have been submitted to the ForCES WG for consideration:

- o FACT
- o GRMP
- o Netlink2

The following sections provide overviews of each of the protocols as well as relevant background information about each protocol.

2.1 FACT

Network Elements (NE) such as routers are becoming more and more complex as they try to cope with demanding features like policy based routing, firewalls and NATs, and QoS aware routing. As a result, issues like scalability, (the ability to cost-effectively grow a network as demand increases) and extensibility (the ability to dynamically configure the network for some specific services by programming the NEs that handle those services) become very important. The ForCEs protocol has been specified to help resolve these issues by decoupling control and forwarding elements of a network element, and also adding extensibility features to the NE.

FACT (Forwarding And Control ElemenT) protocol has been designed for exchanging information between control elements (CEs) and forwarding elements (FEs) distributed in a ForCES network element (NE). The relationship between CEs and FEs is a master/slave one. The FACT protocol is logically separated into a base protocol and an extensible data model defined in [9]. It consists of a common, fixed size header and variable size payload which carries the information defined by the data model. All FACT messages are 32-bit aligned.

FACT's messages are grouped into six (6) classes namely:

1) Connection and Association messages, which help establish logical connections between FEs and CEs,
2) Capabilities Control messages, which the CE uses to query and configure the capabilities of the FE,
3) State Maintenance messages, which are used to track element states,
4) Traffic Maintenance messages, which are used exchanging control packets between CEs and FEs,
5) Event Notification messages used for reporting asynchronous events, and
6) Vendor Specific messages which are used to extend the protocol beyond its current capabilities.

FACT supports versioning and priority, and its unique design of separating control and data traffic into different channels helps reduce the threat of Denial of Service (DoS) attacks making the protocol more robust. It provides reliability by using a reliable transport protocol, thus simplifying the protocol design. It also provides failover mechanisms that can exploit redundant elements in the system or network element.

The FACT protocol follows the basic design principles of simplicity, reuse of existing mechanisms and enabling easy interoperability. In this respect, FACT reuses existing transport and security protocols

which are widely available and avoids building such mechanisms into

the protocol which can increase complexity. It also mandates single transport, security mechanisms and payload encapsulation which help with enabling easy interoperability. The clean separation of FACT base protocol and data model also helps with simplicity and extensibility.

2.2 GRMP

General Router Management Protocol (GRMP) Version 1 intends to be as a ForCES protocol, which acts at the Fp reference point in the ForCES framework. GRMP is designed to meet all the requirements for the ForCES protocol at the Fp reference point.

GRMP protocol is a master-slave protocol. CEs act as masters and FEs as slaves. Slaves have rights to send to masters request, response or report messages, while masters can send command messages to slaves as well as send request, response, or report messages. GRMP protocol acts in a mode of a base-protocol associated with a data model, where GRMP is as a base-protocol and ForCES FE model as a Data Model. GRMP defines basic management messages, while managed data are defined in the associated ForCES FE model. Most of the data types and functional descriptions related to specific IP services such as routing service conforming to RFC 1812, QoS configurations, high-touch capabilities like NAT and firewall should be expressed by Logical functional Blocks (LFBs) and LFB topologies. The ForCES protocol application layer is responsible on how to configure the LFBs and the LFB topologies based on the FE capabilities in order to implement specific IP services and QoS resources deployment.

GRMP is developed separately from General Switch Management Protocol (GSMP) protocol. However, GRMP has been considering its possible compatibility with GSMP.

GRMP protocol is composed of protocol messages. GRMP organizes these messages according to the different object types and layers in ForCES architecture the protocol intends to manage, as follows:

1. FE Coarse Layer
. FE management messages
These messages take a whole FE as the managed object. Messages of this type include that for operation of FE join or leave, FE action, FE attribute, FE event report, etc. Messages of this type also include that for GRMP slave management, which is a module GRMP protocol has defined in a FE that is responsible for protocol message interpreting, executing, generating, and encapsulating at the FE side.

2. FE Fine Layer

. LFB management messages

This type of messages is for the management of LFB layer operation. It takes LFBs as its managed objects. Messages of this type include that for operation of LFB action, LFB attribute, etc.

. Datapath management messages
This type of messages is for the management of datapaths in an FE. It takes datapathes as the managed objects.

3. CE Layer
. CE Informing messages
This type of messages takes CE as the operated object. Because CE acts as a master in ForCES protocol, allowed operations to CE from FE are only that like CE attribute query, CE event report, etc.

4. Protocol Layer and others
Messages of this type include:
. GRMP ACK message
. Packet redirection messages
. GRMP Batch messages
. Managed Object(MO) management messages
In order to support network management tools like SNMP in ForCES architecture, GRMP provides these management messages. The messages take Managed Objects (MOs) defined in some specific network management tools as their operating objects. Operations of MOs are that like MO get, MO set, and MO response.

From the perspective of the message communication in between CE and FE, GRMP messages can be divided into following types:

1. Messages for query and response types. These messages can be from CE to FE, or from FE to CE.

2. Messages for command and configuration types. These messages are only from CE to FE.

3. Messages for report types. These messages can be from CE to FE, or from FE to CE.

GRMP has defined a "Object Class" prefix [3 Section 3.4.5] to allow managed objects to be defined inside GRMP protocol, by different versions of ForCES FE models, or by vendors, in order to make GRMP protocol more scalable and flexible regarding its managed entities.

2.3 Netlink2

Netlink2 [4] is a proposal for the ForCES post-association phase protocol. It is derived from Linux Netlink [10], and as such it builds on the experience gained by use of Netlink for forwarding and

control separation in thousands of Linux-based NEs such as routers, security gateways, NATs, bridges, etc. Netlink2 has incorporated extensions to Netlink to allow multi-host distributed operation across local and global networks.

The key features of Netlink2 are the following:

- Peer-to-peer protocol which can be used between an arbitrarily large set of addressable elements (CEs or FEs).

- Embedded addressing of source and destination addressable elements.

- Support for unicast, logical, and broadcast addressing.

- Separation from the underlying transport protocol. Netlink2 can run directly over unreliable IP/UDP, or can run over TCP/IP or SCTP/IP. It can also run directly over native link-layer protocols (e.g., Ethernet, PCI).

- Application-layer support for synchronization, sequencing, and acknowledgement (not dependent on the underlying transport) provides for element association, reliable or unreliable message exchanges, and atomic transactions.

- Congestion control by means of underlying transport protocol or by means of Netlink2 flow control between addressable entities.

- Support for message priority.

- Support for message bundling and fragmentation.

- Simultaneous support for unicast and multicast communication. Multiple Netlink2 wires can be established between CEs and FEs for efficient message exchange. Multicast wires can enhance scalability in certain local circumstances.

- Flexible ACK strategy support to minimize multicast ACK implosion.

- Support for FE_instance:LFB_instance and FE_group:LFB_class addressing.

- Support for a variety of command/query modes.

- Authentication support by means of option headers.

- Support for protocol versioning and extension headers.

3. Architectural Requirements Compliance Evaluation

This section contains a review of each protocol proposal's level of compliance to the ForCES architecture requirements. Many of the architectural requirements will be instantiated in some fashion in the protocol selected. Given that the architectural requirements are not direct protocol requirements, the review below will consist of prose rather than specific levels of compliance as is used in the protocol section below.

3.1 FACT

FACT fulfills all the protocol requirements listed in section 5. By doing this it in turn supports all the architectural requirements defined in the ForCES Requirements [5]. FACT supports the separation of the NE into CE and FE components, with CE handling roles such as control, signaling and routing data calculation. The CE configures the FE with all the information necessary for the FE's proper operation. The FE's functions could be layer-3 forwarding, NAT, metering, shaping, firewall, etc. Also, FACT state maintenance messages help resolve the various states of the distributed CEs and FEs to provide a unified state of the NE.

3.2 GRMP

GRMP protocol is designed based on the ForCES architecture requirements. We review its compliance to the individual requirement items as below:

1) For architecture requirement #1
GRMP packets can be transported via any suitable mediums, such as TCP/IP, Ethernet, ATM fabrics, and bus backplanes.

2) For architecture requirement #2
ForCES requires that FEs MUST support a minimal set of capabilities necessary for establishing network connectivity (e.g., interface discovery, port up/down functions). This process is usually out of the range of the ForCES protocol, but GRMP protocol has no restriction on this functionality.

3) For architecture requirement #3
By properly configuring FEs with their LFBs in a NE via GRMP protocol, packets can arrive at one FE and depart at the other FE or FEs. In the case where more than one CE work simultaneously in a NE, the consistency and synchronization of control of the CEs is basically required, but which is beyond the scope of the ForCES protocol.

4) For architecture requirement #4

By properly configuring LFBs in FEs in a NE via GRMP protocol, the NE can appear as a single functional device in a network. In the case more than one CE work simultaneously in a NE, the consistency and synchronization for the CEs to control FEs is basically required, but this is beyond the scope of the ForCES protocol.

5) For architecture requirement #5
ForCES protocol requirement #2 has comprised this architecture requirement, refer to Section 5.2.2 for details on GRMP compliance to this requirement.

6) For architecture requirement #6
Please refer to Section 5.13.2 for details.

7) For architecture requirement #7
Please refer to Section 5.8.2 for details.

8) For architecture requirement #8
Please refer to Section 5.9.2 for details.

9) For architecture requirement #9
GRMP supports RFC1812 compliant router functions by means of following mechanisms in GRMP:
-Fully supporting ForCES FE model
-Packet redirection messages
-Datapath management messages
-Managed Object(MO) management messages

10) For architecture requirement #10
In GRMP, FE topology query and response messages [3 Section 4.1.3] are used for CEs to query FE topology information in a NE.

11) For architecture requirement #11
Please refer to Section 5.3.2 for details.

12) For architecture requirement #12
Please refer to Section 5.11.2 for details.

13) For architecture requirement #13
GRMP supports multiple FEs working together in a NE by using FE identifiers and by allowing CEs to be informed of FE topology information. GRMP supports multiple CEs working together in a NE by supporting CE redundancy or failover functionality.

14) For architecture requirement #14
GRMP defines Managed Object (MO) management messages [3 Section 4.5] to meet the requirement.

A MO is an object defined by some network management tool, such as the object defined by Object Identifier in SNMP MIBs. MO management messages work in the way as below:

1. Query of MOs resident in an FE can be directly implemented by network management tools.
2. Change of MOs resident in an FE can only be made via a CE. To do this, the high touch LFBs in the FE will redirect all network management protocol messages like SNMP messages concerning MO changes to the CE, then the CE will use the MO management messages described in this section to change values of MOs in the FE. Of course, if necessary, query of the MOs can also be made via the CE.
3. MOs resident in a CE can be directly queried or changed by the CE with CE high touch capability. Before the CE can do this, network management messages still need to be redirected from FEs to the CE.

3.3 Netlink2

Section 5 of [5] identifies a set of ForCES architecture requirements, some of which have an impact on the design and features of the ForCES post-association phase protocol. The following items document the compliance of the Netlink2 proposal [4] to the each of the ForCES architectural requirements.

1. Netlink2 is capable of operating over IP, therefore it is capable of operating over any link-layer technology supporting IP. Netlink2 is also capable of operating directly over link-layer technologies in the absence of IP. This is made possible due to the inclusion of the following protocol mechanisms:

   - Source and Destination element addresses.
   - Message priority
   - Message length
   - Sequence number
   - ACK/NACK support
   - Checksum option (available in the Netlink2 extension header)

   Note that in the case of both IP and direct link-layer operation, Netlink2 depends on the pre-association protocol to associate Netlink2 CE/FE addresses to link-layer (and, if applicable, IP) addresses.

2. Not applicable.

3. Netlink2 supports generic unicast transmission to/from any particular FE.

4. Netlink2 supports a state machine and the necessary protocol messages to support the transition from pre-association to post-association phase.

5. Netlink2 supports CE and FE authentication by means of authentication and name protocol options. The authentication mechanism design is not documented in [4]. The qualified name mechanism is intended to be derived from [11].

6. Netlink2 supports autonomous message generation by FEs.

7. Netlink2 supports NOOP messages which can be used to implement a heartbeat protocol between addressable elements. One component of CE redundancy (state synchronization) is enabled by allowing secondary CEs to be participants in Netlink2 multicast wires.

8. Netlink2 allows multiple unicast and/or multicast wires to be established between CEs and FEs, allowing separate channels for data and control exchanges. Configuration of the necessary LFBs for enabling packet redirection is opaque to Netlink2.

9. Netlink2 messages can be addressed to individual LFBs on an FE (or to all LFBs of a class on all FEs within a particular multicast group). The configuration commands for specific FE LFBs is opaque to Netlink2. Command templates derived from [10] are documented in [4] and may be used with Netlink2.

10. FE topology information may be conveyed by Netlink2 (but is not defined by it).

11. Netlink2 includes a variety of mechanisms to enhance scalability. Message exchanges (e.g., physical interface configuration) specific to a particular FE can be unicast to/from that FE. Message exchanges applicable to a set of multiple FEs (e.g., nexthop updates) can be multicast to all FEs within the set. Netlink2 messages include sufficient addressing information to allow recipients on a multicast wire to quickly filter out messages not intended for them. Partial ACK support is provided to allow reliable multicast communication without ACK implosion at the message originator. ACKs/NACKs for multicast messages can be returned on a unicast wire to the message originator.

12. Netlink2 supports a state machine and the necessary protocol messages for CEs and FEs to join and leave association dynamically.

13. Netlink2 allows support for up to 64K addressable elements (using the currently specified addressing structure).

14.Not applicable.

4. Model Requirements Compliance Evaluation

This section contains a review of each protocol's level of compliance to the ForCES model requirements. The ForCES model will indirectly relate to the protocol in that the protocol will be used to carry information that the model represents. Given that the model requirements are only indirectly related to the protocol selection, the review below will consist of prose rather than specific levels of compliance as is used in the protocol section below.

4.1 FACT

The FACT protocol is logically separated into a base protocol and an extensible payload which can be used to carry the FE, Logical Functional Block (LFB) specific data which is defined by the FE Model [9]. Thus the FACT protocol is cleanly separated from the data model that it carries. The FE Model draft [9] defines the data model for the Forwarding Element and meets all the Model requirements.

FACT's Configure Request and Configure Response message types under the Capabilities Control message group provide a flexible way to configure the functionality of the FE according to the FE Model [9]. The specific parameters needed to assign functionalities and behaviors to the Logical Functional Blocks (LFBs) in the FEs are dictated by the FE Model.

Vendor Specific functions are supported by VS-Data request and VS-Data response messages in the Vendor Specific message group.

4.2 GRMP

GRMP protocol is designed to use ForCES FE model as a base data model for the protocol functionality. GRMP aims to support all operations to all elements defined in ForCES FE model. Following elements for ForCES FE model (including capability model and state model) with their operations are presented in current version of GRMP document:
-FE capabilities
-FE attributes, including FE statistics
-FE events
-LFBs with their attributes (including capabilities, statistics, etc), their actions, and their topologies
-Datapaths
Section 5.1.2 has described GRMP support for the management of the modeled elements. Along with the progress in FE model work, a modification of GRMP can be made to coordinate with the modification in the FE model.

GRMP protocol supports ForCES FE model to meet following model requirements without any restriction from the protocol:

1. Types of logical functions
2. Variations of logical functions
3. Ordering of logical functions
4. Flexibility
5. Minimal set of logical functions

4.3 Netlink2

The Forces FE Information Model [9] will define schemas for describing the capabilities and attributes of FEs and LFBs. From these, protocol TLVs will be derived. These TLVs will be communicated as the payload of ForCES protocol messages. The payload of Netlink2 messages is opaque to Netlink2, with the exception that the Netlink2 header includes a message type field which can be used to convey information about the content of the message payload (this feature could be ignored by defining a generic NLMSG_FECMD type). Netlink2 supports message addressing at the granularity of FE_instance:LFB_instance or FE_group:LFB_class. Further, it supports a variety of flag fields (request, root, match, atomic, replace, exclusive, create, and append) which support efficient atomic transaction, configuration, and query exchanges as may be required by the FE model.

5. Protocol Requirements Compliance Evaluation

This section contains a review of each protocol's level of compliance to the ForCES protocol requirements. Given that the protocol requirements are directly related to the protocol proposals, a very concrete method is used in reviewing compliance - the following key identifies the level of compliance for each of the following protocols to each protocol requirement in the ForCES requirements RFC:

T = Total compliance. Meets the requirement fully.

P+ = Partial compliance. Fundamentally meets the requirement through the use of extensions (e.g. packages, additional parameters, etc.)

P = Partial compliance. Meets some aspect of the requirement, however, the necessary changes require more than an extension and/or are inconsistent with the design intent of the protocol.

N = Not compliant. Does not meet the requirement.

Each subsection of this section begins with the specific protocol requirement text found in the ForCES requirements.

5.1 Protocol Requirement: Configuration of Modeled Elements

The ForCES protocol MUST allow the CEs to determine the capabilities of each FE. These capabilities SHALL be expressed using the FE model whose requirements are defined in Section 6. Furthermore, the protocol MUST provide a means for the CEs to control all the FE capabilities that are discovered through the FE model. The protocol MUST be able to add/remove classification/action entries, set/delete parameters, query statistics, and register for and receive events.

5.1.1 FACT

FACT's Capabilities Control message class contains Configure Request and Configure Response messages that can be used to configure the FE's behavior from the CE. Also, the Capability request and response messages can be used by the CE to query and learn the FE capabilities. Please see section 5.2 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.1.2 GRMP

Most of GRMP protocol messages are for the management of modeled elements in ForCES FEs. They are listed as follows:

1) FE capability query and response messages [3 Section 4.1.4]
2) FE attribute manipulate message [3 Section 4.1.6]
3) FE attribute query and response messages [3 Section 4.1.7]
4) FE event report message [3 Section 4.1.8]
5) LFB action manipulate message [3 Section 4.2.1].
6) LFB topology query and response messages [3 Section 4.2.2]
7) LFB attribute manipulate message [3 Section 4.2.3].
8) LFB attribute query and response messages [3 Section 4.2.4]
9) Datapath Manipulate Message [3 Section 4.3.1]
10) Datapath query and response messages [3 Section 4.3.2]

Protocol requirement compliance level: (T)

5.1.3 Netlink2

Netlink2 includes service template definitions that allow modeled elements to be configured using TLVs. As the model gets refined, appropriate modifications to those TLVs can be made without modifying the Netlink2 base protocol.

Protocol requirement compliance level: (T)

5.2 Protocol Requirement: Support for Secure Communication

a) FE configuration will contain information critical to the functioning of a network (e.g. IP Forwarding Tables). As such, it MUST be possible to ensure the integrity of all ForCES protocol messages and protect against man-in-the-middle attacks.
b) FE configuration information may also contain information derived from business relationships (e.g. service level agreements). Because of the confidential nature of the information, it MUST be possible to secure (make private) all ForCES protocol messages.
c) In order to ensure that authorized CEs and FEs are participating in a NE and defend against CE or FE impersonation attacks, the ForCES architecture MUST select a means of authentication for CEs and FEs.
d) In some deployments ForCES is expected to be deployed between CEs and FEs connected to each other inside a box over a backplane, where physical security of the box ensures that man-in-the-middle, snooping, and impersonation attacks are not possible. In such scenarios the ForCES architecture MAY rely on the physical security of the box to defend against these attacks and protocol mechanisms May be turned off.
e) In the case when CEs and FEs are connected over a network, security mechanisms MUST be specified or selected that protect the ForCES protocol against such attacks. Any security solution used for ForCES MUST specify how it deals with such attacks.

5.2.1 FACT

FACT uses TLS when its endpoints are running over an IP network or in an insecure environment. For a closed box or physically secure environment, it is possible to turn off the protocol security functions. The security association between the CEs and FEs is established before any FACT association establishment messages are exchanged. Also, FACT recommends using rate limiting mechanisms on the FE to protect against DoS attacks. Please see section 8 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.2.2 GRMP

1) When GRMP messages are encapsulated in a IP based medium, GRMP protocol recommends to use IPsec or TLS [3 Section 4.1.2] to authenticate the CEs and FEs and to secure the communication between CEs and FEs to defend against possible man-in-the-middle or replay attacks. GRMP has no restrictions on using other approaches for secure communications. When GRMP messages are transported over bus backplanes or in the case CEs and FEs are physically all in one box,

the secure mechanism to defend man-in-the-middle attach MAY be turned off.

2) [3 Section 4.6] has addressed the GRMP mechanism to prevent DoS attacks.

3) [3 Section 4.1.2] has addressed the method to prevent possible FE join or leave flood attacks.

Protocol requirement compliance level: (T)

5.2.3 Netlink2

Secure communication is supported at the Netlink2-wire level using pre-configured mechanisms (TLS, IP-SEC, MSEC, etc), in the case security cannot be achieved by physical means only. Netlink2 introduces ForCES qualified names (fqn) that permit the authentication of FEs and CEs based on names instead of potentially variable addresses. The definition of fqns remains to be completed.

Protocol requirement compliance level: (P+)

5.3 Protocol Requirement: Scalability

The ForCES protocol MUST be capable of supporting (i.e., must scale to) at least hundreds of FEs and tens of thousands of ports. For example, the ForCES protocol field sizes corresponding to FE or port numbers SHALL be large enough to support the minimum required numbers. This requirement does not relate to the performance of a NE as the number of FEs or ports in the NE grows.

5.3.1 FACT

FACT can support up to 64K FEs and 64K CEs at the same time due its 16 bit addressing range of both the CE-Tag and FE-Identifier fields. Please see section 4.1 in [2] for more details on this. In addition, it uses TCP (for IP interconnection between CEs and FEs) which provides congestion control and thus helps in supporting the scalability requirement.

Protocol requirement compliance level: (T)

5.3.2 GRMP

In GRMP, a FE is identified by a 16 bits FE Identifier [3 section 3.2], which is theoretically able to identify up to 64k FEs.

Possible limitation in GRMP protocol to FE port number may be from FE port address space, maximum number of list elements in "list data

format" [3 section 3.4.3], and LFB instance identifier space. The evaluations of scalability for them are as follows:

1) An Addressable Entity (AE) address data format is defined in GRMP [GRMP_3.4.4], which is theoretically capable of describing any length of addresses of AEs, therefore FE port address space is not limited.

2) Element number of a list in "list data format" [3 Section 3.4.3] is expressed with 16 bits data space, which theoretically limits list element number within 64k.

3) LFB instance ID [3 Section 4.2.1] is expressed using 16 bits data space, which can also theoretically represent 64k instances of one kind of LFB such as a port LFB.

Protocol requirement compliance level: (T)

5.3.3 Netlink2

Netlink2 includes a flexible multicast-capable addressing mechanism (32-bits). This allows it to take full advantage of wires capable of supporting multicast/broadcast, such as IP multicast-based wires or Ethernet multicast-based wires (for the local scope environment). In addition, Netlink2 does not limit the number and types of wires that can be used (instead of a single pair of unicast-based control and data channels). Netlink2 wires are configured during pre-association. Support for multicast at the ForCES level is key for scalable distribution (in terms of CPU usage and bandwidth) of identical routing tables from a CE to multiple FEs, for instance. Note that depending on the available transport mechanisms (or lack thereof), a ForCES multicast wire may be implemented using suboptimal multi-unicast TCP connections, for instance.

Protocol requirement compliance level: (T)

5.4 Protocol Requirement: Multihop

When the CEs and FEs are separated beyond a single hop, the ForCES protocol will make use of an existing RFC2914 compliant L4 protocol with adequate reliability, security and congestion control (e.g. TCP, SCTP) for transport purposes.

5.4.1 FACT

FACT uses TCP as the transport protocol which is congestion aware and meets the transport requirements for multi-hop IP networks. Please see section 3.2 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.4.2 GRMP

GRMP aims to be capable of supporting remote control that allows CEs and FEs to separate multihops away, as well as supporting close or very close proximity control of CEs and FEs. When the CEs and FEs are separated beyond a single hop, GRMP RECOMMENDS using an RFC2914 compliant L4 protocol such as TCP, SCTP for the protocol message transmission with adequate reliability, security and congestion control [3 Section 3.3].

Protocol requirement compliance level: (T)

5.4.3 Netlink2

Congestion control and flow control may be necessary depending on the scope in which the ForCES protocol operates. Congestion control is particularly relevant in the global scope, and can be provided by the transport mechanisms used for the Netlink2 wires. Flow control may be provided either by the transport mechanism, by means of backpressure (such as local scope case with a switching fabric interconnecting CEs and FEs) or by an appropriate windowing of Netlink2 messages. Netlink2 accomodates multi-hop wires (i.e., global scope) using any appropriate congestion-control-friendly transport protocol, such as TCP or SCTP. At a minimum, Netlink2 requires that UDP is available for the local scope, and TCP (congestion control and reliability) and/or SCTP-PR (congestion control and timeliness)
(Note: decision should be taken by the working group) for the global scope.

Protocol requirement compliance level: (T)

5.5 Protocol Requirement: Message Priority

The ForCES protocol MUST provide a means to express the protocol message priorities.

5.5.1 FACT

FACT supports up to 8 levels of priority using the 3 priority bits in the common header. Please see section 4.1.6 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.5.2 GRMP

GRMP defines a priority field at GRMP message header [3 Section 3.2] to express the protocol message priority.

Protocol requirement compliance level: (T)

5.5.3 Netlink2

Netlink2 supports a priority bit in the Netlink2 message header flags, as well as a 16-bit priority field using the Netlink2 header-extension TLV.

Protocol requirement compliance level: (T)

5.6 Protocol Requirement: Reliability

a) The ForCES protocol will be used to transport information that requires varying levels of reliability. By strict or robust reliability in this requirement we mean, no losses, no corruption, no re-ordering of information being transported and delivery in a timely fashion.
b) Some information or payloads, such as redirected packets or packet sampling, may not require robust reliability (can tolerate some degree of losses). For information of this sort, ForCES MUST NOT be restricted to strict reliability.
c) Payloads such as configuration information, e.g. ACLs, FIB entries, or FE capability information (described in section 7, (1)) are mission critical and must be delivered in a robust reliable fashion. Thus, for information of this sort, ForCES MUST either provide built-in protocol mechanisms or use a reliable transport protocol for achieving robust/strict reliability.
d) Some information or payloads, such as heartbeat packets that may be used to detect loss of association between CE and FEs (see section 7, (8)), may prefer timeliness over reliable delivery. For information of this sort, ForCES MUST NOT be restricted to strict reliability.
e) When ForCES is carried over multi-hop IP networks, it is a requirement that ForCES MUST use a RFC 2914 [12]-compliant transport protocol.
f) In cases where ForCES is not running over an IP network such as an Ethernet or cell fabric between CE and FE, then reliability still MUST be provided when carrying critical information of the types specified in (c) above, either by the underlying link/network/transport layers or by built-in protocol mechanisms.

5.6.1 FACT

FACT uses a reliable transport protocol to meet all the reliability requirements. For IP-interconnection between the protocol elements, FACT uses TCP as the transport protocol for the control channel.

Please see section 3.2 in [2] for more details on this. FACT also provides protocol level responses or acknowledgements (and sequence numbers) for control messages.

Protocol requirement compliance level: (T)

5.6.2 GRMP

GRMP supplies two levels of built-in error control mechanisms and several other mechanisms to improve the protocol reliability:

1) Normal level error control
In this level, GRMP protocol uses a specific GRMP ACK message [3 Section 3.4.1] associated with "Result" and "Code" fields in the message headers to protect against errors that may result from message transmission, message processing, or message generating.

2) Strengthened level error control
If higher level of reliability is required for some protocol messages, a built-in error control based on CRC-32 checksums can furthermore be applied [3 Section 3.2]. This makes GRMP able to be transported over some mediums that themselves cannot supply error controls, like Ethernet, UDP, etc.

3) Transaction identifier to control the order of messages
GRMP has defined different transaction identifiers for CE generated messages and for FE generated messages [3 Section 3.2]. This makes it possible to use protocol built-in method to order back protocol messages if in occasional cases messages are reordered.

4) GRMP recommends to use an RFC2914 compliant L4 protocol for message transmission to improve the protocol reliability when CEs and FEs are multihops away [3 Section 3.3].

Protocol requirement compliance level: (T)

5.6.3 Netlink2

Netlink2 defines application-level ACKs to acknowedge that transactions have completed successfully. Reliability can be built using such ACKs only, or can be enhanced using reliable transport protocols (expected to be necessary in the global scope), if available. Each Netlink2 message carries a sequence number as well as a flag that indicates whether an ACK is expected.

Protocol requirement compliance level: (T)

5.7 Protocol Requirement: Interconnect Independence

The ForCES protocol MUST support a variety of interconnect technologies. (refer to section 5, requirement# 1)

5.7.1 FACT

FACT uses interconnect independent addressing (FE Identifier, CE tag) in its common header to provide interconnect independence. For non-IP interconnects, such as ATM, an interconnect specific encapsulation will have to be defined to carry the FACT messages. For IP interconnects, FACT uses TCP as the transport protocol. For non-IP interconnects, which do not provide reliability, the interconnect specific encapsulation might consist of an optional checksum and any other fields to help build reliability, although reuse of existing transport mechanisms is recommended. Please see section 3.1 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.7.2 GRMP

GRMP packets can be transported via any suitable mediums, such as TCP/IP, Ethernet, ATM fabrics, and bus backplanes [3 Section 3.3].

Protocol requirement compliance level: (T)

5.7.3 Netlink2

Netlink2 defines its own addressing. Encapsulations for various non-IP media remain to be defined.

Protocol requirement compliance level: (T)

5.8 Protocol Requirement: CE Redundancy or CE Failover

The ForCES protocol MUST support mechanisms for CE redundancy or CE failover. This includes the ability for CEs and FEs to determine when there is a loss of association between them, ability to restore association and efficient state (re)synchronization mechanisms. This also includes the ability to preset the actions an FE will take in reaction to loss of association to its CE e.g., whether the FE will continue to forward packets or whether it will halt operations. (refer to section 5, requirement# 7)

5.8.1 FACT

FACT exchanges CE and FE element states using the PE State Maintenance messages. FACT also uses Heart-Beat messages (section 5.3 in [2]) to detect protocol element (CE or FE) failure or loss of association between elements and to trigger a switch-over to a

functioning redundant element (CE or FE). Please see section 7.3 in [2] for more details on the different mechanisms (Strong consistency, weak consistency) used for CE failover.

Protocol requirement compliance level: (T)

5.8.2 GRMP

GRMP meets ForCES CE redundancy or CE failover requirement by means of following mechanisms:

1) CE failover or leave policy [3 Section 4.6.4]
This policy is defined as a FE attribute. In this attribute, selectable FE policies for CE failover such as FE graceful restart and CE re-association policies are defined.

2) FE heartbeat policy [3 Section 4.6.6]
The ability to determine the loss of association between a CE and a FE is obtained by use of this FE heartbeat policy and the associated CE heartbeat event.
3) CE status event report [3 Section 4.4.2]

Protocol requirement compliance level: (T)

5.8.3 Netlink2

Netlink2 accommodates transparent CE (and FE) redundancy and failover using Netlink2 multicast wires that include both the active and backup CE (and FE). Using the ECHO flag in the Netlink2 header, a hearbeat mechanism can be created to detect when failover must take place. Actions that take place after a loss of association remains to be defined.

Protocol requirement compliance level: (P+)

5.9 Protocol Requirement: Packet Redirection/Mirroring

a) The ForCES protocol MUST define a way to redirect packets from the FE to the CE and vice-versa. Packet redirection terminates any further processing of the redirected packet at the FE.
b) The ForCES protocol MUST define a way to mirror packets from the FE to the CE. Mirroring allows the packet duplicated by the FE at the mirroring point to be sent to the CE while the original packet continues to be processed by the FE.

Examples of packets that may be redirected or mirrored include control packets (such as RIP, OSPF messages) addressed to the interfaces or any other relevant packets (such as those with Router Alert Option set). The ForCES protocol MUST also define a way for the

CE to configure the behavior of a) and b) (above), to specify which packets are affected by each.

5.9.1 FACT

FACT's Traffic Maintenance Message class includes Control Packet Redirect and Control Packet Forward messages to achieve packet redirection/mirroring. These messages are sent over the separate data channel. Please see section 5.4 in [2] for more details on this. Also, the Event Register/Deregister messages (section 5.5 in [2]) can be used to specify which packets should be redirected/mirrored.

Protocol requirement compliance level: (T)

5.9.2 GRMP

GRMP supports packet redirection by packet redirection messages [3 Section 4.7]. A LFB within LFB topology in a FE should be used to pick out packets that are to be redirected. Packets to be redirected are first put in GRMP slave [3 Section 4.6.1] and then are directed to a CE via the packet redirection message. The attribute of this filter LFB are set by CEs, therefore the CE has the ability to control which packets can be redirected.

To redirect packets from CE to FE, CE just needs to encapsulate the packet to the packet redirection message and send it to the FE. On the FE side, GRMP salve resolves the redirected packet and put it into a datapath in a FE LFB topology so that they can further be delivered by the FE.

By properly configuring LFBs in FE, a packet can be mirrored to CE instead of purely redirected to CE, i.e., the packet is duplicated and one is redirected to CE and the other continues its way in the LFB topology.

Protocol requirement compliance level: (T)

5.9.3 Netlink2

It is expected that the necessary LFB for packet redirection and mirroring is defined by the model itself. The message format to carry redirected packets between the FE and CE remains to be defined.

Protocol requirement compliance level: (P+)

5.10 Protocol Requirement: Topology Exchange

The ForCES protocol MUST allow the FEs to provide their topology information (topology by which the FEs in the NE are connected) to the CE(s). (refer to section 5, requirement# 10)

5.10.1 FACT

FACT's Capabilities and Control Message class includes Query request and response messages to achieve topology information exchange between the CE and FEs. Please see sections 5.2.5, 5.2.6 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.10.2 GRMP

GRMP FE topology query and response messages [3 Section 4.1.3] are used for CEs to query FE topology information in the NE.

Protocol requirement compliance level: (T)

5.10.3 Netlink2

This is expected to be defined by the FE model, therefore it opaque to Netlink2.

Protocol requirement compliance level: (P+)

5.11 Protocol Requirement: Dynamic Association

The ForCES protocol MUST allow CEs and FEs to join and leave a NE dynamically. (refer to section 5, requirement# 12)

5.11.1 FACT

FACT's Connection and Association message class includes Join request, Join response, Leave request and Leave response messages to enable dynamic joining and leaving of protocol elements (CEs, FEs) in the NE. Please see sections 5.1.1, 5.1.2, 5.1.3, 5.1.4 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.11.2 GRMP

In GRMP, specific FE join request message [3 Section 4.1.1] and FE leave request message [3 Section 4.1.2] make FEs able to dynamically join or leave a ForCES NE. While CE failover or leave policy [3 Section 4.6.4] defines the way for CEs to dynamically join or leave

the NE. GRMP also defines FE failover and rejoin policy [3 Section 4.6.5] for FEs to dynamically rejoin the NE.

Protocol requirement compliance level: (T)

5.11.3 Netlink2

Netlink2 uses SYN and FIN messages similarly to TCP to set up and tear down associations. Such messages are sent by default on a broadcast wire, or on pre-configured wires.

Protocol requirement compliance level: (T)

5.12 Protocol Requirement: Command Bundling

The ForCES protocol MUST be able to group an ordered set of commands to a FE. Each such group of commands SHOULD be sent to the FE in as few messages as possible. Furthermore, the protocol MUST support the ability to specify if a command group MUST have all-or-nothing semantics.

5.12.1 FACT

FACT supports command bundling by using multiple TLVs in its message payload. For example, each TLV used in the Configure Request message could represent a different command such as Add, Delete, etc. In addition, FACT also supports 2-phase commit operations. Please see sections 5.2.3, 4.2 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.12.2 GRMP

GRMP supports ForCES protocol command bundling by use of GRMP batch messages [3 Section 4.8]. The messages allow GRMP application layers to pack several different sub message bodies into one single GRMP message. The sub messages are defined to be executed in an all-or-nothing mode.

Protocol requirement compliance level: (T)

5.12.3 Netlink2

Netlink2 supports the concatenation of multiple commands of an identical type in the same Netlink2 message (such as multiple route additions), as well as the bundling of different commands sent in separate Netlink2 messages (using the MULTI flag). All-or-nothing (2-phase commit) is supported using the ATOMIC flag.

Protocol requirement compliance level: (T)

5.13 Protocol Requirement: Asynchronous Event Notification

The ForCES protocol MUST be able to asynchronously notify the CE of events on the FE such as failures or change in available resources or capabilities. (refer to section 5, requirement# 6)

5.13.1 FACT

FACT's Event Notification message class includes the Asynchronous FE Event notification message used to report asynchronous FE events to the CE. Please see section 5.5 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.13.2 GRMP

In GRMP, a FE asynchronously informs CEs of a failure, resources and capabilities changes, and other asynchronous events via GRMP FE event report message [3 Section 4.1.8].

Protocol requirement compliance level: (T)

5.13.3 Netlink2

Netlink2 is peer-to-peer, so any addressable entity can send a message to any other. FE and LFB-level events will have to be defined in the FE model, so that Netlink2 can transmit them.

Protocol requirement compliance level: (T)

5.14 Protocol Requirement: Query Statistics

The ForCES protocol MUST provide a means for the CE to be able to query statistics (monitor performance) from the FE.

5.14.1 FACT

FACT's Capabilities and Control message class includes the Query request and response messages which can be used by the CE for querying the FE's properties and statistics. Please see sections 5.2.5, 5.2.6 in [2] for more details on this.

Protocol requirement compliance level: (T)

5.14.2 GRMP

GRMP defines statistics regarding FE performance as FE or LFB attributes. GRMP uses FE attribute query and response messages [3 Section 4.1.7] and LFB attribute query and response messages [3 Section 4.2.4] to query the statistics.

GRMP can also support query of statistics defined by network management tools like SNMP by using MO get message [3 Section 4.5.1] and MO response message [3 Section 4.5.3].

Protocol requirement compliance level: (T)

5.14.3 Netlink2

Statistics are specific to LFBs and therefore remain opaque to the Netlink2 protocol.

Protocol requirement compliance level: (T)

5.15 Protocol Requirement: Protection Against Denial of Service Attacks

Systems utilizing the ForCES protocol can be attacked using denial of service attacks based on CPU overload or queue overflow. The ForCES protocol could be exploited by such attacks to cause the CE to become unable to control the FE or appropriately communicate with other routers and systems. The ForCES protocol MUST therefore provide mechanisms for controlling FE capabilities that can be used to protect against such attacks. FE capabilities that MUST be manipulated via ForCES include the ability to install classifiers and filters to detect and drop attack packets, as well as to be able to install rate limiters that limit the rate of packets which appear to be valid but may be part of an attack (e.g. bogus BGP packets).

5.15.1 FACT

FACT uses separate control and data channels to provide robustness in the protocol against Denial of Service (DoS) attacks. Please see section 3.3 in [2] for more details on this. Also, the Configure Request and Response messages in FACT could be used to install filters on FEs which can be used for rate-limiting the malicious traffic.

Protocol requirement compliance level: (T)

5.15.2 GRMP

GRMP supports protection against DoS attacks by means of following mechanisms:

1) A model for GRMP slave module [3 Section 4.6.1]

In this model, all GRMP messages sending to CE are put into two different channels: the data channel, which is only for packet redirection messages, and the control channel, which is for other GRMP messages. Messages on the two channels pass through a packet scheduler for the link connecting to CE. The scheduler is managed by CE by setting some scheduling policies (disciplines) to it. In this way, the CE can control the traffic over the two channels dynamically according to the monitored traffic status, to defend against DoS attacks and to protect control channel transmission.

2) GRMP DoS protection policy [3 Section 4.6.2]
In this policy, scheduling priorities, channel bandwidths, and congestion control policies for the individual data channel and control channel can be set.

3) GRMP DoS attack alert policy [3 Section 4.6.3]

4) A DoS attack alert event report [3 Section 4.1.8]

Protocol requirement compliance level: (T)

5.15.3 Netlink2

Netlink2 defines Netlink2 SYN Cookies as a mechanism to prevent DoS attacks originating in a environment where security cannot be physically ensured. Netlink2 relies on appropriate policers to rate limit data traffic redirected to CEs. As different wires may be used for data and control traffic, prioritization and reliability/unreliability can be chosen appropriately for each wire with a suitable transport protocol.

Protocol requirement compliance level: (T)

5.16 Protocol Requirement Summary Table

This section is a summary of the compliance levels claimed for each protocol above and is included as a convenience.

Protocol Requirement	FACT	GRMP	Netlink2
1. Configuration of Modeled Elements	T	T	T
2. Support for Secure Communication	T	T	P+
3. Scalability	T	T	T
4. Multihop	T	T	T
5. Message Priority	T	T	T
6. Reliability	T	T	T
7. Interconnect Independence	T	T	T
8. CE Redundancy or CE Failover	T	T	P+
9. Packet Redirection/Mirroring	T	T	P+
10. Topology Exchange	T	T	P+
11. Dynamic Association	T	T	T
12. Command Bundling	T	T	T
13. Asynchronous Event Notification	T	T	T
14. Query Statistics	T	T	T
15. Protection Against Denial of Service Attacks	T	T	T

Security Considerations

This document is a comparison between three protocols in order to help in the selection of the best approach to use as the ForCES protocol. Security considerations are addressed in each of the protocol proposals and MUST be included as part of the fitness evaluation for each proposal.

References

1 Bradner, S., "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

2 Audu, A. et al., "ForwArding and Control ElemenT protocol (FACT)", work in progress, November 2003, <draft-gopal-forces-fact-06.txt>

3 Wang, W. et al., "General Router Management Protocol (GRMP) Version 1", November 2003, <draft-wang-forces-grmp-01.txt>

4 Salim, J. H. et al., "Netlink2 as ForCES Protocol", work in progress, October 2003, <draft-jhsrha-forces-netlink2-02.txt>

5 Khosravi, H. et al., "Requirements for Separation of IP Control and Forwarding", RFC 3654, July 2003.

6 Yang, L. et al., "Forwarding and Control Element Separation (ForCES) Framework", work in progress, October 2003, <draft-ietf-forces-framework-10.txt>

7 Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997

8 Barnes, M., "Middlebox Communications (MIDCOM) Protocol Evaluation", work in progress, Nov 2002, <draft-ietf-midcom-protocol-eval-06.txt>

9 Yang, L. et al., "ForCES Forwarding Element Functional Model", work in progress, October 2003, <draft-ietf-forces-model-01.txt>

10 Hadi Salim, et. al., "Linux Netlink as an IP Services Protocol", RFC 3549, July 2003.

11 M. Bakke, et. al., "iSCSI Naming and Discover", work in Progress, June 20, 2003, <draft-ietf-ips-iscsi-name-disc-10.txt>

12 S. Floyd, "Congestion Control Principles", RFC2914, September 2000.

Author's Addresses

Alex Audu
Alcatel R&I
1000 Coit Road
Plano, TX 75075
U.S.A.
Phone: 1-972-477-7809
Email: alex.audu@alcatel.com

Steven Blake
Ericsson
920 Main Campus Drive, Suite 500
Raleigh, NC 27606
U.S.A.
Email: steven.blake@ericsson.com

Ligang Dong
Hangzhou University of Commerce
149 Jiaogong Road
Hangzhou, 310035, P.R.China
Phone: +86-571-88071024
Email: donglg@mail.hzic.edu.cn

Robert Haas
IBM Research
Zurich Research Laboratory
Saeumerstrasse 4
CH-8803 Rueschlikon,
Switzerland
Email: rha@zurich.ibm.com

Hormuzd Khosravi
Intel
2111 NE 25th Avenue
Hillsboro, OR 97124
U.S.A.
Phone: 1-503-264-0334
Email: hormuzd.m.khosravi@intel.com

David Putzolu
Intel
Mailstop JF3-206-H10
2111 NE 25th Avenue
Hillsboro, OR 97124
U.S.A.
Phone: 1-503-264-4510
Email: david.putzolu@intel.com

Jamal Hadi Salim
Znyx Networks
195 Stafford Rd. West
Ottawa, Ontario
Canada
Email: hadi@znyx.com

Weiming Wang
Department of Information and Electronic Engineering
Hangzhou University of Commerce
149 Jiaogong Road
Hangzhou, 310035, P.R.China
Phone: +86-571-88057712
Email: wangwm@hzcnc.com